A NEW METHOD OF
MULTIDIMENSIONAL SCALING

by

James Robert Capra

United States
Naval Postgraduate School

THESIS

A NEW METHOD OF MULTIDIMENSIONAL SCALING

by

James Robert Capra

June 1970

A NEW METHOD OF MULTIDIMENSIONAL SCALING

by

James Robert Capra
Lieutenant (junior grade), United States Naval Reserve
B.A., Georgetown University, 1968

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 1970

ABSTRACT

This paper is based on the multidimensional scaling technique of Joseph B. Kruskal. It is comprised of three parts: The first part describes Kruskal's objectives and introduces his goodness of fit measure, called stress; the second part discusses some problems associated with Kruskal's technique, focusing on the concept of stress; in the third part, an alternate goodness of fit measure, called V, is proposed, together with a different procedure for doing multidimensional scaling. Part three also includes a discussion of the superiority of V over stress as a goodness of fit measure.

TABLE OF CONTENTS

I. MULTIDIMENSIONAL SCALING

Like all statistical techniques, multidimensional scaling is a method of summarizing and drawing inferences from a large body of data. In this case, the data are the judgments made by a respondent about the similarities or differences between stimuli presented in pairs. For N stimuli, multidimensional scaling attempts to find N points in a t dimensional mapping whose interpoint distances (N(N-1)/2 of them in all) somehow resemble or match the corresponding N(N-1)/2 similarity-dissimilarity judgments made by the respondent.

The importance of the number _t_ stems from its interpretation as the number of dimensions on which the respondent based his judgments. The best method for determining this number when the investigator is using the multidimensional scaling techniques to be discussed in this paper has been given by Joseph B. Kruskal. (Kruskal, 1964a) His method assumes the capability to derive a mapping for any number of dimensions (one, two, three or more) and then involves a comparison of these mappings of different dimensionality. Since the question of how to derive a mapping for an arbitrary number of dimensions is the main topic of this paper, the dimensionality of the mapping which multidimensional scaling seeks to derive will be two throughout this paper. The techniques for deriving a mapping are the same whether the dimensionality is one, two, three or more. Also, the mapping will always be in Euclidean space. The contents of this paper can be adapted with very little trouble, however, to non-Euclidean spaces based on a city-block metric or a Minkowski r metric. (Kruskal, 1964a)

The discussion can be simplified by the use of an example. Suppose one is interested in identifying the dimensions of appeal of political candidates. What factors make some candidates attractive to a respondent and other candidates unattractive? For simplicity, suppose the investigator examines the feelings of one respondent with respect to four political candidates. Multidimensional scaling would help the investigator determine these factors or dimensions of appeal by providing him with a t (two in this case) dimensional mapping of the candidates. The mapping would be based on judgments made by the respondent about the similarities or differences between the candidates presented in pairs.

One method of eliciting the judgments of a respondent concerning the similarities or differences between candidates presented in pairs is to administer a simple questionnaire to him. A typical item in such a questionnaire might resemble the following:

> Please specify how similar or how different these two individuals are in their general appeal to you by circling one of the numbers, 1 through 9. If you circle number 1, it implies that they are exactly equal in their general appeal to you, while if you circle number 9, it implies that they are extremely different in their general appeal to you.

		Exactly Equal								Extremely Different
1.	Lyndon B. Johnson Hubert H. Humphrey	1	2	3	4	5	6	7	8	9

If the respondent's feelings toward four candidates were to be examined, he would be asked the same question about 5 other pairs of candidates, making a total of 6 questions in all. (4(3)/2)

The basic premise underlying the analysis of data from a questionnaire of this kind is that the numbers circled are measures of psychological distance, closeness or proximity between stimuli for the respondent. Shepard calls them proximity measures. (Shepard, 1962a) Here, however, they will be called psychological distances. These psychological distances will be labeled δ_{ij}'s, with the i referring to one stimulus and the j referring to the other. The investigator only obtains N(N-1)/2 judgments from the respondent since δ_{ij} equals δ_{ji} by assumption, and a special experimental design is required if δ_{ii} is to have any meaning. (If the assumption were dropped and the special design employed, the method of analysis would not change.) The formula N(N-1)/2 can be obtained by counting the elements in the lower triangular portion of an N by N matrix or by using the formula for the number of combinations of N objects taken two at a time, which is $\binom{N}{2}$ or N(N-1)/2.

A number of computer-based procedures for doing multidimensional scaling are currently available. (Shepard 1962, Kruskal 1964, Lingoes 1965) However, the discussion in this paper will be limited to the most popular of these, the procedure proposed by Joseph B. Kruskal in 1964. In addition to being the most widely used, Kruskal's technique is the best vehicle for the introduction of a slightly different technique in this paper. For the most part, Kruskal's notation will be used in the analysis to follow.

Since the properties of the δ_{ij}'s will become important later, it should be noted that they are measurements on an ordinal scale. In other words, the investigator can say that a δ_{ij} of 8 is greater than one of 5; however, the difference between the two (for example, 3) is not meaningful. The latter property accompanies both linear interval and ratio scales, but not an ordinal scale. To obtain interval proximity measures or psychological distances (δ_{ij}'s), one would need an experimental model somewhat different from the one outlined by Kruskal and used in this paper. For example, interval measures can be obtained by the "method of multidimensional rank order," the "method of complete triads," or a number of other methods. (Torgerson 1958) All of these methods are based on the law of comparative judgment. It should be noted, however, that even the law of comparative judgment does not yield δ_{ij}'s that are measurements on a ratio scale, a point that will become important later. (Thurstone 1920)

As mentioned earlier, the investigator has obtained N(N-1)/2 distance judgments from the respondent. Let M equal N(N-1)/2. These psychological distances, δ_{ij}'s, have a certain rank order:

$$\delta_{i_1 j_1} < \delta_{i_2 j_2} < \ldots\ldots\ldots < \delta_{i_m j_m} < \ldots < \delta_{i_M j_M} .$$

For example, a respondent might provide the following answers to a four candidate questionnaire:

$$\delta_{12}=6 \qquad \delta_{14}=9 \qquad \delta_{34}=1$$
$$\delta_{13}=8 \qquad \delta_{24}=7 \qquad \delta_{23}=2 \quad .$$

This would mean that

$$\delta_{34}<\delta_{23}<\delta_{12}<\delta_{24}<\delta_{13}<\delta_{14} \quad .$$

Multidimensional scaling seeks to obtain a two (or t) dimensional mapping, called a configuration, of the stimuli for which the Euclidean (or non-Euclidean if they are desired) distances between the stimuli have the same rank order as the psychological distances, or δ_{ij}'s. This is the isomorphism which multidimensional scaling seeks to create between the psychological distances or proximity measures and the inter-point distances in a Euclidean mapping. Let X_i be a two dimensional vector, x_{i1} and x_{i2}, referring to the ith political candidate's position in the two dimensional mapping in Euclidean space. The Euclidean distance between the two candidates, i and j, is the square root of the sum of squares of the distances along each axis, or by the Pythagorean theorem,

$$d_{ij} = \sum_{t=1}^{2} (x_{it}-x_{jt})^2 \quad .$$

In the four candidate example, the investigator would want to find a two dimensional mapping of the candidates for which $d_{34} \leq d_{23} \leq d_{12} \leq d_{24} \leq d_{13} \leq d_{14}$. The only fixed characteristics of the mapping are the relationships between the d_{ij}'s. The axes can be rotated in any direction and the origin placed anywhere. Kruskal places the origin at the centroid of the configuration and normalizes the configuration by making the sum of the squared distances of the points from the origin

equal one. Finally, he "normalizes the angular attitude of the configuration by rotating it so that its so-called principal axes coincide with the coordinate axes (in the natural order)."[1] The principal axes rotation is very important in the achievement of a solution for a different multidimensional scaling technique, that of Roger N. Shepard.[2] However, it is not important for solution purposes in the Kruskal technique, although it might help the investigator in the interpretation of his results.

Of course not all configurations of the points (particular mappings of the candidates) will yield d_{ij}'s that have the same rank order as the δ_{ij}'s. Consequently, what the investigator needs and what Kruskal provides is an index to determine how close a given configuration comes to satisfying the rank order requirements which the δ_{ij}'s place on the d_{ij}'s. This index is called stress.

Prior to defining stress, Kruskal introduces a new set of symbols, called $\hat{d}_{ij}$'s. The $\hat{d}_{ij}$'s are numbers which completely satisfy the rank order requirements given by the δ_{ij}'s. If the d_{ij}'s themselves satisfy these requirements, then the

---

[1]Kruskal, Joseph B., "Nonmetric Multidimensional Scaling: A Numerical Method," Psychometrika, v. 29, p. 120, June 1964.

[2]Shepard, Roger N., "The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function," Psychometrika, v. 27, p. 132, June 1962.

set of $\hat{d}_{ij}$'s could be, and in fact will be, identical to the set of d_{ij}'s. However, consider the following situation. The δ_{ij}'s are in the order specified in the example used earlier,

$$\delta_{34}<\delta_{23}<\delta_{12}<\delta_{24}<\delta_{13}<\delta_{14} \quad ,$$

and the mapping that has been obtained has the following d_{ij}'s:

$d_{34}=2$	$d_{12}=3$	$d_{13}=7$
$d_{23}=1$	$d_{24}=4$	$d_{14}=6$

The rank order of the d_{ij}'s is the following:

$$d_{23}\leq d_{34}\leq d_{12}\leq d_{24}\leq d_{14}\leq d_{13} \quad .$$

A set of numbers, $\hat{d}_{ij}$'s, that satisfy the rank order constraints set by the δ_{ij}'s can be obtained in the following way:

$$\hat{d}_{34}=\hat{d}_{23}=(d_{34}+d_{23})/2=1.5 \qquad \hat{d}_{24}=d_{24}$$

$$\hat{d}_{12}=d_{12} \qquad \hat{d}_{13}=\hat{d}_{14}=(d_{13}+d_{14})/2=6.5 \quad ,$$

so that $\hat{d}_{34}\leq\hat{d}_{23}\leq\hat{d}_{12}\leq\hat{d}_{24}\leq\hat{d}_{13}\leq\hat{d}_{14}$.

This example demonstrates that the $\hat{d}_{ij}$'s are based on <u>averages of certain d_{ij}'s.</u> In the example, so-called "equality blocks" (for lack of a better name) were created for $\hat{d}_{34}$ and $\hat{d}_{23}$ and for $\hat{d}_{13}$ and $\hat{d}_{14}$ by averaging d_{34} and d_{23} to find $\hat{d}_{34}$ $(=\hat{d}_{23})$ and averaging d_{13} and d_{14} to find $\hat{d}_{13}$ $(=\hat{d}_{14})$. The method of calculation of $\hat{d}_{ij}$'s for every situation is part of a technique called "monotone regression." (Miles 1959) Monotone regression is not discussed in any detail in this paper. However, one of its properties is that the differences

between the d_{ij}'s and the $\hat{d}_{ij}$'s computed in the example represent the minimum differences between the distances, the d_{ij}'s, and any set of numbers satisfying the rank ordering specified by the δ_{ij}'s.

In the above paragraph the point was made that if the d_{ij}'s do not satisfy the rank order constraints, the $\hat{d}_{ij}$'s will be averages of certain d_{ij}'s, as seen in the example. If the problem has M distances, then it can be shown that there are $2^{M-1}-1$ possible ways to average the d_{ij}'s to obtain $\hat{d}_{ij}$'s; or, if the case under which each $\hat{d}_{ij}$ equals its respective d_{ij} is considered to be a degenerate type of averaging, then 2^{M-1} possible ways exist.[3]

Another example may help. Suppose the investigator is dealing with three stimuli and consequently with three distances: d_{12}, d_{13}, d_{23}. The psychological distances are in the following order: $\delta_{12}<\delta_{13}<\delta_{23}$. There are 2^{3-1} or 4 different ways to average d_{ij}'s to obtain $\hat{d}_{ij}$'s. First of all, each $\hat{d}_{ij}$ may be equal to its respective d_{ij}, or

$$\begin{aligned} &\hat{d}_{12}=d_{12} \\ (1)\qquad &\hat{d}_{13}=d_{13} \\ &\hat{d}_{23}=d_{23} \end{aligned}$$

[3]The proof of this statement is a lengthy one that must be performed inductively. Since the number 2^{M-1} is not crucial to this analysis, the proof will not be given here.

Another possibility is that

$$(2) \qquad \begin{aligned} \hat{d}_{12}=\hat{d}_{13}&=(d_{12}+d_{13})/2 \\ \hat{d}_{23}&=d_{23} \end{aligned} \quad .$$

A third is that

$$(3) \qquad \begin{aligned} \hat{d}_{12}&=d_{12} \\ \hat{d}_{13}=\hat{d}_{23}&=(d_{13}+d_{23})/2 \end{aligned} \quad .$$

The final possibility is that

$$(4) \qquad \hat{d}_{12}=\hat{d}_{13}=\hat{d}_{23}=(d_{12}+d_{13}+d_{23})/3 \quad .$$

Monotone regression would lead to one of the four specifications, depending on the order of the d_{ij}'s obtained from a particular mapping. For example, given that $\delta_{12}<\delta_{13}<\delta_{23}$, the second specification would be appropriate if

$$\begin{aligned} &d_{12}>d_{13} \\ &d_{12}\leq d_{23} \\ &d_{13}<d_{23} \end{aligned} \quad .$$

Each of the four specifications will be called a block equality system. In the fourth specification, the block equality is $\hat{d}_{12}=\hat{d}_{13}=\hat{d}_{23}$, by definition. In the third, $\hat{d}_{13}$ equals $\hat{d}_{23}$ by definition, while in the second specification, $\hat{d}_{12}$ equals $\hat{d}_{23}$ by definition. There are no defined equalities in the first specification.

Now that the method of obtaining the $\hat{d}_{ij}$'s from the d_{ij}'s has been outlined and the concept of a block equality system as a defined equality between $\hat{d}_{ij}$'s has been introduced,

stress can be defined:

$$\text{Stress} = \sqrt{\frac{\sum_{m=1}^{M} (d_{i_m j_m} - \hat{d}_{i_m j_m})^2}{\sum_{m=1}^{M} d_{i_m j_m}^2}}$$

The heart of Kruskal's technique is the derivation of the points (the X's) in the mapping, and subsequently their distances. Nonlinear programming becomes relevant at this point since the problem is to find the points and their distances that do the following:

Minimize Stress

Subject to:

$$\hat{d}_{i_1 j_1} \leq \hat{d}_{i_2 j_2} \leq \ldots \leq \hat{d}_{i_m j_m} \leq \ldots \leq \hat{d}_{i_M j_M} \quad .$$

Kruskal employs the "method of steepest descent" to solve this problem. (Kruskal 1964b) His use of this method implies that he is treating the minimization as an unconstrained one, since this method is generally employed in unconstrained minimization problems. (Spang 1962) As Spang points out, the use of the "method of steepest descent" for constrained minimization problems, which Kruskal in fact does, requires the construction of a Lagrangian and then the unconstrained minimization of the Lagrangian. Kruskal uses the "method of steepest descent" but mentions neither Lagrangians nor the convexity assumptions that are normally made when minimizing a Lagrangian. It should also be noted, in passing,

that the formula for the gradient on page 125 (Kruskal 1964) is incorrect since it fails to take into account the fact that the $\hat{d}_{ij}$'s change as the d_{ij}'s change.

A more conventional nonlinear programming approach to this problem shows that Kruskal's technique actually derives a solution to one of 2^{M-1} different constrained minimizations (nonlinear programming problems). There is one nonlinear programming problem for each different block equality system or definition of the $\hat{d}_{ij}$'s.[4] The "method of steepest descent" leads to the solution to one of these 2^{M-1} different problems. However, the "method" by itself cannot determine if the solution to another of these 2^{M-1} problems (where the $\hat{d}_{ij}$'s are defined differently) would have a lower stress value than the one which it has derived. There could be $2^{M}-1$ other minima, under different block equality systems, that are smaller than the one yielded by the "method of steepest descent." (Of

---

[4]Define, a priori, the relationship between the $\hat{d}_{ij}$'s and and the d_{ij}'s; then stress becomes a function of the d_{ij}'s alone (the X's or points in the mapping, ultimately). The constraints represented by $\hat{d}_{i_1j_1} \leq \hat{d}_{i_2j_2} \ldots \leq \hat{d}_{i_mj_m} \ldots \leq \hat{d}_{i_Mj_M}$ are really constraints on the d_{ij}'s, since the relationship between the d_{ij}'s and the $\hat{d}_{ij}$'s has been specified beforehand. The problem of minimizing stress, then, has been transformed into a conventional nonlinear programming problem. However, there exist 2^{M-1} possible relationships between the d_{ij}'s and the d_{ij}'s (block equality systems) and, consequently, 2^{M-1} nonlinear programming problems. Of course, some problems may not have solutions since in some cases the constraints may imply a feasible region that is the null set. Spang (1962) contains a discussion of many techniques that could be employed to solve these constrained minimizations. Since the publication of that article, other techniques have been developed that might prove helpful. (Klingman 1963, Glass and Cooper 1965, Box 1965)

course, there could also be none.) A different approach to the problem that would bypass the above difficulty might be possible.[5] However, as will be shown in the next part, in most cases, stress is not a good index of goodness of fit in the first place. Even if a better minimization technique were possible, it would have no effect upon the suitability of stress as a measure of goodness of fit. Consequently, the next section will be devoted to a discussion of some of the problems inherent in the concept of stress.

---

[5]There appears to be a certain ordering to the 2^{M-1} different problems in the sense that stress is always lower when the $\hat{d}_{ij}$'s are defined in one way than when they are defined in another way. For example, stress is always lower when each $\hat{d}_{ij}$ is defined to be equal to its respective d_{ij} than when the $\hat{d}_{ij}$'s are defined in any other way. If this ordering could be determined, then the first nonlinear programming problem that had a feasible solution would be the one having the lowest minimum. The ordering may only be a partial one, however, which would complicate things considerably.

II. STRESS AS A MEASURE OF GOODNESS OF FIT

A number of problems with Kruskal's goodness of fit measure, stress, become evident upon closer examination. One of these is the question of the meaning of stress, which, in turn is related to the problem of the specification of both minimum and maximum possible values that the index can attain. Another problem is the question of whether or not stress, which will be shown to be dependent on ratios between the d_{ij}'s, is an appropriate measure of the goodness of fit of the rank order of the d_{ij}'s to the rank order dictated by the δ_{ij}'s. These problems of interpretation, maximum and minimum possible values and appropriateness of stress will be discussed in that order in this second part.

The meaning of stress is the first question to be raised in this part. The square of stress would appear to lend itself to interpretation as the percentage of the variance of the d_{ij}'s not conforming to the monotonic (rank order) requirements set by the δ_{ij}'s. Under this interpretation, stress itself (not stress squared) would then be the square root of this number or the percentage of the standard deviation of the d_{ij}'s not accounted for by the monotonic requirements.

This interpretation of stress encounters problems as soon as one examines the maximum and minimum possible values of stress. Intuitively, the minimum should be 0.0 and the maximum 1.0. Intuition is only partially correct in this case.

Obviously, if the d_{ij}'s perfectly satisfied the monotonic requirements, then stress would be 0.0, since each $\hat{d}_{ij}$ would

equal its respective d_{ij}, as discussed earlier. However, the conditions under which stress would be equal to 1.0, which would be the logical maximum under the "percentage of variance" interpretation, are not clearly defined.

Stress would be 1.0 if all $\hat{d}_{ij}$'s were zero. However, since the d_{ij}'s by definition cannot be less than zero and all are not allowed to equal zero at the same time, a degenerate solution which Kruskal disallows,[6] all the $\hat{d}_{ij}$'s cannot equal zero. Similar problems arise if one tries to make each $\hat{d}_{ij}$ equal to twice its respective d_{ij}, the other condition under which stress would equal 1.0.

The following is a short proof that in a particular problem it is impossible for stress to equal 1.0. (Under the "percentage of variance" interpretation, it should always be possible for stress to equal 1.0.) Once again, suppose that an investigator is using only three stimuli and the specified rank ordering of the distances is as before:

$$\delta_{12}<\delta_{13}<\delta_{23}$$

As mentioned earlier, there are four possible block equality systems for this problem:

$$(1) \qquad \begin{aligned} \hat{d}_{12}&=d_{12} \\ \hat{d}_{13}&=d_{13} \\ \hat{d}_{23}&=d_{23} \end{aligned}$$

---

[6]Kruskal, <u>op. cit.</u>, p. 120.

(2) $$\hat{d}_{12}=\hat{d}_{13}=(d_{12}+d_{13})/2$$
$$\hat{d}_{23}=d_{23}$$

(3) $$\hat{d}_{12}=d_{12}$$
$$\hat{d}_{13}=\hat{d}_{23}=(d_{13}+d_{23})/2$$

(4) $$\hat{d}_{12}=\hat{d}_{13}=\hat{d}_{23}=(d_{12}+d_{13}+d_{23})/3 \quad .$$

Stress will then equal

$$\sqrt{\frac{(d_{12}-\hat{d}_{12})^2+(d_{13}-\hat{d}_{13})^2+(d_{23}-\hat{d}_{23})^2}{d_{12}^2+d_{13}^2+d_{23}^2}} \quad .$$

If the first block equality system is used and a mapping that allows $\hat{d}_{12}\leq\hat{d}_{13}\leq\hat{d}_{23}$ is obtained, stress would equal 0.0. Consequently, if stress were to equal 1.0, it would do so under one of the other three block equality systems. Assume that stress can equal 1.0 under block equality system number two. Then the following relationships must hold:

$$\frac{(d_{12}-(d_{12}+d_{13})/2)^2+(d_{13}-(d_{12}+d_{13})/2)^2+(d_{13}-d_{13})^2}{d_{12}^2+d_{13}^2+d_{23}^2} = 1.0$$

$$\frac{1}{4}((d_{12}-d_{13})^2+(d_{13}-d_{12})^2 = d_{12}^2+d_{13}^2+d_{23}^2$$

$$2d_{12}^2-4d_{12}d_{13}+2d_{13}^2 = 4d_{12}^2+4d_{13}^2+4d_{23}^2$$

$$-4d_{12}d_{13} = 2d_{12}^2+2d_{13}^2+4d_{23}^2 \quad .$$

The last relationship is an obvious contradiction since the right side of the equation must be greater than the left

side ($d_{ij} \geq 0$) unless all d_{ij}'s are equal to zero (which is not allowed). The same type of contradiction arises when block equality system number three is employed. Under block equality system number four, the following "illegal" statement is obtained:

$$-2d_{12}d_{13}-2d_{12}d_{23}-2d_{13}d_{23}=d_{12}^2+d_{13}^2+d_{23}^2 \quad .$$

The lack of a clearly defined maximum of 1.0 for stress makes a "percentage of variance" interpretation difficult at best. However, these problems are not nearly as serious as those associated with the question of the appropriateness of stress as a measure of goodness of fit. This question is very closely related to the discussion in the first part about levels of measurement. The reader may want to refer to that section at this time.

Euclidean distances are numbers on a ratio scale. Both the differences between two distances and the ratios between two distances are meaningful. As mentioned earlier, the proximity measures or psychological distances will normally be measurements on an ordinal scale, although they may be measurements on a linear interval scale if the law of comparative judgment is invoked. Whether the psychological distances are ordinal or interval measurements, the ratios between two δ_{ij}'s are not meaningful.

The major problem with stress as a measure of goodness of fit of the d_{ij}'s to the δ_{ij}'s is that it depends on the

ratio properties of the d_{ij}'s. The following example will demonstrate that stress can be reduced to a function of the ratios between d_{ij}'s.

The same three candidate example will be used. Assume, once again that the respondent has specified that $\delta_{12}<\delta_{13}<\delta_{23}$. Further, suppose a mapping which has the following distance relationships has been obtained:

$$d_{12}>d_{13}$$

$$d_{12}\leq d_{23}$$

$$d_{13}<d_{23} \quad .$$

In order to insure that $\hat{d}_{12}\leq\hat{d}_{13}\leq\hat{d}_{23}$, block equality system number two must be used, or:

$$\hat{d}_{12}=\hat{d}_{13}=(d_{12}+d_{13})/2$$

$$\hat{d}_{23}=d_{23} \quad .$$

Under these conditions,

$$\text{Stress} = \sqrt{\frac{(d_{12}-(d_{12}+d_{13}))^2+(d_{13}-(d_{12}+d_{13}))^2}{d_{12}^{\;2}+d_{13}^{\;2}+d_{23}^{\;2}}} \quad ,$$

or,

$$\text{Stress}^2 = \frac{((d_{12}-d_{13})/2)^2+((d_{13}-d_{12})/2)^2}{d_{12}^{\;2}+d_{13}^{\;2}+d_{23}^{\;2}} \quad .$$

Let d_{13}/d_{12} equal K and let d_{23}/d_{12} equal G. Then,

$$\text{Stress}^2 = \frac{K^2-2K+1}{2+2K^2+2G^2} \quad ,$$

or stress is entirely dependent on the ratios d_{23}/d_{12} and d_{13}/d_{12}.

The problem is obvious. If the investigator is using ordinal δ_{ij}'s, stress is supposedly a measure of how well the d_{ij}'s fit the rank ordering specified by the respondent, that is the rank ordering of the δ_{ij}'s. This would seem to indicate that it should not be dependent on a property which the original data do not possess, that is the property that the ratios of the distances are meaningful. Notice the effect of G^2 in the above equation. Stress decreases as G^2 increases. (Recall that G equals d_{23}/d_{12}.) The respondent might have specified that $\delta_{23}=6$ and $\delta_{12}=5$. Why should one obtain a lower stress value when $d_{23}=1000$ and $d_{12}=2$ than when $d_{23}=3$ and $d_{12}=2$?

When the δ_{ij}'s are interval measurements, that is when the law of comparative judgment has been invoked, the same problem arises. Why should the measure of goodness of fit be dependent on the ratios between the d_{ij}'s when the ratios between the δ_{ij}'s are not meaningful?

An example of what can happen when stress is used as a measure of goodness of fit may help. Suppose an investigator wants to examine a 4 candidate situation and the respondent specifies that $\delta_{34}<\delta_{23}<\delta_{12}<\delta_{24}<\delta_{13}<\delta_{14}$. First, suppose he has obtained a mapping with the following d_{ij}'s,

$d_{34}=1$	$d_{12}=3$	$d_{13}=10$
$d_{23}=2$	$d_{24}=4$	$d_{14}=8$.

Stress in this situation equals $\sqrt{\frac{2}{194}}$. Now suppose he obtains another mapping with the following set of d_{ij}'s:

$d_{34}=2$ $\quad$ $d_{12}=3$ $\quad$ $d_{13}=10$

$d_{23}=1$ $\quad$ $d_{34}=4$ $\quad$ $d_{14}=9$.

For this second mapping, stress is equal to $\sqrt{\frac{1}{211}}$.

Notice that the first mapping violated the rank ordering specified by the δ_{ij}'s only once while the second mapping violated the rank ordering twice. Yet the first had a higher stress value than the second

In the next part, a new index of goodness of fit will be proposed for the case or ordinal δ_{ij}'s, and a very simple method of dealing with interval δ_{ij}'s will be discussed.

III. HOW TO FIT EUCLIDEAN DISTANCES TO PSYCHOLOGICAL DISTANCES

In this part a new index of goodness of fit, V, will be introduced. This index is sensitive to neither the size of the difference between two d_{ij}'s nor to the size of the ratio between two d_{ij}'s. It also has a well-defined maximum of 1.0 and a well-defined minimum of 0.0. This simple index is based on the number of violations of the order relations specified by the δ_{ij}'s. A very simple method of dealing with interval δ_{ij}'s will also be discussed in this section.

Earlier, the point was made that if one were attempting to map N stimuli into Euclidean space, then there would be N(N-1)/2 psychological distances, δ_{ij}'s, associated with these stimuli, and likewise N(N-1)/2 Euclidean distances, d_{ij}'s, associated with the mapping. Again, let M=N(N-1)/2. The respondent, by his answers, specifies a rank ordering for the δ_{ij}'s:

$$\delta_{i_1 j_1} < \delta_{i_2 j_2} \ldots < \delta_{i_m j_m} < \ldots < \delta_{i_M j_M} .$$

The problem of multidimensional scaling is to find a mapping of the N stimuli. The Euclidean distances between the points (stimuli) in the mapping should have, as nearly as possible, the same rank order as that of the psychological distances. Implicit in this rank order are M(M-1)/2 constraints:

$$d_{i_1 j_1} \leq d_{i_2 j_2}$$

$$d_{i_1 j_1} \leq d_{i_3 j_3}$$

$$\vdots$$

$$d_{i_1 j_1} \leq d_{i_m j_m}$$

$$\vdots$$

$$d_{i_1 j_1} \leq d_{i_M j_M}$$

$$\vdots$$

$$d_{i_m j_m} \leq d_{i_{m+1} j_{m+1}}$$

$$\vdots$$

$$d_{i_m j_m} \leq d_{i_M j_M}$$

$$\vdots$$

$$d_{i_{M-1} j_{M-1}} \leq d_{i_M j_M}$$

For any particular mapping, then, a possible index of the goodness of fit of the mapping to the rank order constraints would be the number of violations of the constraints by the d_{ij}'s. In fact, this is the index that will be adopted, except for one obvious alteration. The number of violations of the

constraints by the d_{ij}'s should be expressed as a percentage of the maximum possible number of violations, or in other words,

$$V = \frac{A}{M(M-1)/2} \quad ,$$

where A is the actual number of violations and M(M-1)/2 is the maximum possible number of violations. If V equals 0.0, no violations occur and the mapping perfectly satisfies the rank order constraints. If V equals 1.0, the mapping perfectly violates the rank order constraints.

If this new index were adopted, the problem of multidimensional scaling would become the problem of finding a configuration that minimizes V. This minimization is very similar to a problem encountered in mathematical programming, the derivation of an initial feasible point. (Klingman 1963, Hilleary 1966, Rosen 1961) A very popular technique for finding a feasible point is Hooke and Jeeves' direct search algorithm for unconstrained functions. (Hilleary 1966, Klingman 1963, Hooke and Jeeves 1961)

The above references contain complete descriptions of the direct search algorithm. In order to use the algorithm to minimize V, one would start with an arbitrary configuration of the N points in t dimensions. The t coordinate values for each point would be the independent variables, making Nt independent variables in all. A univariate search is first performed, with each independent variable being changed by a

small amount, one at a time, in order to determine the direction toward the minimum. If this exploratory move succeeds in lowering the objective function, a "pattern move" is then attempted. A pattern move is a move based on the directions of the last two (sometimes more than two) exploratory moves. Various modifications of the algorithm tend to differ with respect to the weights given to previously successful exploratory moves. If the pattern move does not succeed in lowering the objective function, another exploratory move is attempted. Eventually, the exploratory move will be unable to lower the objective function. In that case, the step size of the search is reduced and another search is performed. The process is repeated until the step size reaches a predetermined minimum.

The new index, V, vears a striking similarity to Kendall's tau, a commonly used rank correlation coefficient. (Kendall 1962) In fact the two indices can be related by the following equation:

$$\text{tau} = 1.0 - 2V \quad .$$

Kendall's tau was not adopted as the index of goodness of fit for two reasons. First, certain characteristics of V are similar to characteristics possessed by Kruskal's index, stress. For example, a perfect fit of the d_{ij}'s to the nonmetric hypothesis would yield a value of 0.0 for both stress and V. For both indices, a low value is interpreted as a good fit while a high value is interpreted as a poor fit. Of course, with Kendall's tau, a perfect fit would yield a value of 1.0. A

certain amount of consistency among indices of goodness of fit seems desirable, and consequently V should be the preferred index on this basis. Also, the "percentage of possible violations" interpretation of V is intuitively appealing.

A second reason for adopting V instead of tau is based on a disadvantage which both possess, but to a different extent. Neither V nor tau is a continuous variable. V is not continuous since the numerator A, is discrete. In any problem, the number of violations can be 0,1,2,3 and so forth up to M(M-1)/2. The difference between successive values of A is 1. Kendall's tau has the same denominator as V; however, the numerator is different. The difference between successive values of the numerator is 2. In other words, V is on a more compressed scale than tau. (The formula relating the two indices also demonstrates this fact.) Kendall has shown that as the denominator in the expression for tau (M(M-1)/2 in this case) becomes large, tau approximates a continuous variable.[7] In fact, he has shown that for a denominator greater than 45 (M greater than 10), tau can be considered to be a continuous variable. The compressed scale of V approaches continuity even faster than tau, since the difference between successive values for the numerator is 1, not 2.

A necessary condition for continuity of the function relating values of the coordinates of the points in a Euclidean

---

[7]Kendall, M. G., Rank Correlation Methods, 3rd ed., p. 69, Charles Griffin and Company, 1962.

mapping and V is that V itself be continuous. Consequently, if the only impediment to the continuity of the function is the lack of continuity of V (which is the case here), then as V approaches (at the limit) a continuous variable, the function will approach a continuous one. Most minimization techniques require the assumption of continuity of the objective function. Consequently, the index which allows the function to approach a continuous one faster (V in this case) should be preferred.[8]

It should be noted that stress is a continuous function. However, as mentioned earlier, the minimization of stress is a minimization under constraints. On the other hand, like the problem of finding a feasible point, the minimization of V is essentially an unconstrained one.[9]

Like Kendall's tau, V is obviously not sensitive to the magnitude of the difference between two d_{ij}'s nor to the size of their ratio. As mentioned earlier, V has both a clearly defined minimum of 0.0 and a clearly defined maximum of 1.0.

---

[8]In all likelihood, if the Hooke and Jeeves technique will work with V as the index it will probably work with tau as the index. In fact, one could probably adopt the Spearman rho as the goodness of fit measure if he desired to do so.

[9]One set of constraints is operative in this problem. The d_{ij}'s are not allowed to equal zero. These constraints can be handled effectively by the insertion of a penalty function into the Hooke and Jeeves algorithm. This function would automatically set the value of V equal to 1.0 when a configuration with one or more zero distances is tested.

Both of these properties contrast markedly with the properties of stress, which is dependent on the ratios between the d_{ij}'s and does not have a clearly defined maximum of 1.0. Also, Kendall has proposed a very simple way of dealing with ties.[10] His method can be used in the computation of V, in the event that certain δ_{ij}'s or certain d_{ij}'s are equal.

If the Hooke and Jeeves' technique, or some other algorithm will in fact minimize V, the index would appear to have another desirable property that Kruskal's stress does not clearly possess. Suppose an investigator were to obtain a mapping of 14 political candidates that minimized V. For interpretation purposes, he might want to examine the constraints that were violated. It might happen that a large portion of the violations (if not all of them) involved a particular stimulus, candidate number 1, for example. (That is, d_{ij} is not less that d_{34} as it should be; neither are d_{13}, d_{14} and so forth.) This kind of result might indicate that the respondent based his judgments about candidates 2 through 14 on two dimensions (for example, liberalism and good looks) while randomly making judgments about candidate 1, or making them on the basis of something other than the dimensions of liberalism and good looks. This might be very important to an investigator who is trying to interpret the dimensions.

---

[10]Kendall, op. cit., pp. 34-48.

One can imagine instances when there might be a number of different mappings that will yield the same minimum V value, but with different constraints being violated. It would appear to be possible to eliminate at least some, if not all, of these mappings from consideration by choosing the one(s) with the smallest number of stimuli involved in violations. In fact, it may even be possible to insert this criterion into the minimization problem.

Now that a new index of goodness of fit when the δ_{ij}'s are ordinal measures has been derived and discussed, it should be clear what kind of index ought to be used when the δ_{ij}'s are interval measures. The easiest index to use would probably be the Pearson r. The problem would be one of seeking the maximum r between the δ_{ij}'s and the d_{ij}'s, or the minimum negative of r. Again, the minimization would be an unconstrained one. A direct search technique could again be used. The "optimum gradient" method or one of the other gradient techniques discussed in Spang (1962) might be more efficient than the Hooke and Jeeves technique in this case, however. Since r is continuous, the continuity problems inherent in the use of V or tau do not arise in this case.

SUMMARY

The formulation of the new measure of goodness of fit, V, and the discussion of the use of the Pearson r for interval data complete the line of argument followed in this paper. In the first part, multidimensional scaling was defined. The ordinal or, under certain conditions, interval nature of the data which are used as input into most multidemensional scaling techniques was discussed. Finally, one approach to multidimensional scaling, that of Joseph B. Kruskal, was discussed in detail. The second part highlighted three problems with Kruskal's measure of goodness of fit. These were the problems of interpretation, a well-defined maximum possible value and appropriateness of the index. Most of the discussion in the third and final part was concerned with a new measure of goodness of fit when the data are measurements on an ordinal scale. The relationship between V and Kendall's tau showed that V is essentially a measure of the rank correlation between the distances, d_{ij}'s, implicit in a particular mapping of the stimuli, and the psychological distances or δ_{ij}'s which an investigator obtains from a respondent. A method of minimizing V was suggested in this part and the problem of continuity was discussed. Among the desirable properties which V possesses are ease of interpretation, clearly defined maximum and minimum possible values and insensitivity to properties of the d_{ij}'s which the δ_{ij}'s do not possess. Finally, a straightforward extension of the use of the rank correlation between the d_{ij}'s and δ_{ij}'s was proposed for the case when the δ_{ij}'s are

measurements on an interval scale, namely the Pearson r or linear correlation coefficient. The next task to be performed, in a subsequent analysis, is, of course, the programming of a technique for minimizing V. After a routine is implemented, the output should be systematically compared to output from Kruskal's routine. Once this task has been completed, the distribution of V under various conditions should be examined carefully. As David Klahr has pointed out (Klahr 1969), this type of analysis alone will allow the investigator to make probability statements about the goodness of fit value which he has obtained.

BIBLIOGRAPHY

Barton, D. E. and Mallows, C. L., "The Randomization Bases of the Amalgamation of Weighted Means," Journal of the Royal Statistical Society, Series B, V. 23, p. 423-433, 1961.

Beals, R., Krantz, D. H., and Tversky, A., "Foundations of Multidimensional Scaling," Psychological Review, V. 75, p. 127-142, June 1968.

Box, M. J., "A New Method of Constrained Optimization and a Comparison With Other Methods," Computer Journal, V. 8, p. 42-52, April 1965.

Fletcher, R. and Powell, M. J. D., "A Rapidly Convergent Descent Method for Minimization," Computer Journal, V. 6, p. 163-168, July 1963.

Glass, H. and Cooper, L., "Sequential Search: A Method for Solving Constrained Optimization Problems," Journal of the Association for Computing Machinery, V. 12, p. 71-82, January 1965.

Naval Postgraduate School Technical Report/Research Paper 59, The Tangent Search Method of Constrained Minimization, by R. R. Hilleary, March 1966.

Hooke, Robert and Jeeves, T. A., "Direct Search Solution of Numerical and Statistical Problems," Journal of the Association for Computing Machinery, V. 8, p. 212-229, April 1961.

Kendall, M. G., Rank Correlation Methods, 3rd ed., Charles Girffin and Company, London, 1962.

Klahr, David, "A Monte Carlo Investigation of the Significance of Kruskal's Nonmetric Scaling Procedure," Psychometrika, V. 34, p. 319-330, September 1969.

Klingman, William Robert, Nonlinear Programming By the Multiple Gradient Summation Technique, Master's Thesis, University of Texas, August 1963.

Klingman, William Robert and Himmelblau, D. M., "Nonlinear Programming With the Aid of the Multiple Gradient Summation Technique," Journal of the Association for Computing Machinery, V. 11, p. 400-415, October 1964.

Kruskal, Joseph B., "Multidimensional Scaling By Optimizing Goodness of Fit to a Nonmetric Hypothesis," Psychometrika, V. 29, p. 1-28, March 1964.

Kruskal, Joseph B., "Nonmetric Multidimensional Scaling: A Numerical Method," Psychometrika, V. 29, p. 28-42, June 1964.

Lingoes, J. C., "An IBM 7090 Program for Guttman-Lingoes Smallest Space Analysis," Behavioral Science, V. 10, p. 183-184, 1965.

Miles, R. E., "The Complete Amalgamation Into Blocks, By Weighted Means, Of a Finite Set Of Real Numbers," Biometrika, V. 46, p. 317-327, 1959.

Rosen, J. B., "The Gradient Projection Method for Nonlinear Programming, Part II. Nonlinear Constraints," S.I.A.M. Journal, V. 9, p. 514-532, 1961.

Shepard, R. N., "The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function, I," Psychometrika, V. 27, p. 125-140, 1962.

______________, "The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function, II," Psychometrika, V. 27, p. 219-264, 1962

Siegel, S., Nonparametric Statistics, McGraw-Hill, 1956.

Spang, H. A., III, "A Review of Minimization Techniques for Nonlinear Functions," S.I.A.M. Review, V. 4, p. 343-365, October 1962.

Thurstone, L. L., "A Law of Comparative Judgment," Psychological Review, V. 34, p. 273-286, 1927.

Torgerson, W. S., Theory and Methods of Scaling, John Wiley and Sons, 1958.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Asst. Professor R. S. Elster, Code 66Ea Department of Business Administration & Economics Naval Postgraduate School Monterey, California 93940	1
4.	LT(j.g.) James R. Capra, USNR 3550 Otis Street Wheat Rige, Colorado 80033	1

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

A New Method of Multidimensional Scaling

4. DESCRIPTIVE NOTES *(Type of report and, inclusive dates)*

Master's Thesis; June 1970

5. AUTHOR(S) *(First name, middle initial, last name)*

James Robert Capra

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1970	38	22

8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

This paper is based on the multidimensional scaling technique of Joseph B. Kruskal. It is comprised of three parts: The first part describes Kruskal's objectives and introduces his goodness of fit measure, called stress; the second part discusses some problems associated with Kruskal's technique, focusing on the concept of stress; in the third part, an alternative goodness of fit measure, called V, is proposed, together with a different procedure for doing multidimensional scaling. Part three also includes a discussion of the superiority of V over stress as a goodness of fit measure.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

Security Classification

A-31408

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Stress						
Multidimensional scaling						
Psychological distance						
Proximity measures						
Rank order correlation						